Visuo-haptic Perception of Objects and Scenes

Article · June 2010

DOI: 10.1007/978-1-4419-5615-6_14

Fiona N Newell
Trinity College Dublin

Chapter 14
Visuo-haptic Perception of Objects and Scenes

Fiona N. Newell

14.1 Introduction

Although both the visual and the tactile modalities can extract, encode, and process spatial information from objects for the purpose of recognition and localisation, relatively little is understood about how such information is shared across these modalities. The aim of this chapter is to provide an overview of our current understanding of multisensory integration for the purpose of high-level perception. This chapter is structured around evidence in support of the idea that vision and touch contribute to two basic perceptual tasks, namely object recognition and the perception of the spatial location of objects, and that these sensory modalities process information for the purpose of recognition or localisation in a very similar manner. Moreover, visual and tactile processing of spatial object information is underpinned by shared neural substrates, although the extent to which these substrates are shared seems to be task dependent. I will argue that it is the common way in which information is processed by vision and touch, underpinned by shared neural substrates, that allows for efficient sharing of information across these modalities for recognition and localisation.

The reader may be interested to note the historical context to the topic of this chapter. The year 2009 marks the tercentenary of the publication of George Berkeley's (1709) *An Essay Towards a New Theory of Vision* [first edition] which he penned whilst he was a scholar in Trinity College Dublin. In his essay he declares that he will "consider the Difference there is betwixt the Ideas of Sight and Touch, and whether there be any Idea common to both Senses". The proposals outlined in his essay offered such a unique insight into the issue of how vision and touch contribute to perception that they still resonate today. Indeed, 300 years following the publication of this essay, many researchers around the world are grappling with the very same issues laid out in Berkeley's essay. On the face of it, this may sound rather

---

F.N. Newell (✉)
School of Psychology and Institute of Neuroscience, Lloyd Building, Trinity College,
Dublin 2, Ireland
e-mail: fiona.newell@tcd.ie

M.J. Naumer, J. Kaiser (eds.), *Multisensory Object Perception in the Primate Brain*,
DOI 10.1007/978-1-4419-5615-6_14,

pessimistic as is suggests that not much progress has been achieved in the science of visuo-tactile integration in 300 years. On the contrary, the current state of knowledge is a significant achievement but before we could provide empirical evidence for the philosophical questions raised by Berkeley it was first necessary for several other significant scientific discoveries to be made. These include (but are most certainly not limited to) Darwin's theory of evolution and the consequent development of comparative studies to provide insight into neuronal processes in multisensory integration; advances in neuroscience such as Cajal's discovery of the synapse and Hubel and Wiesel's discovery of the structure of the visual cortex; the emergence of new scientific disciplines such as experimental psychology and cognitive neuroscience; and the advent of technology such as computers and neuroimaging. These and many other advances have afforded us a multidisciplinary account of how the senses contribute to and result in a coherent perception of the world around us.

To highlight the advances in our understanding particularly of how vision and touch share information about objects and the layout of objects that surround us, research studies have generally focused on either the perception of object information for recognition or the perception of space for localisation. Over the following sections of this chapter, I will review to what extent information is shared across these modalities for the recognition of objects and spatial arrangements of objects in scenes and provide evidence for common functional organisation of these modalities for the purpose of each task.

14.2 Evidence for Common Principles of Functional Organisation Across Vision and Touch

Over the past few decades, research into visual processing has provided evidence that this system is structurally and functionally separated into two streams, namely the occipitotemporal (i.e. "what") and occipitoparietal (i.e. "where") stream. Accordingly, each stream is involved in the processing of visual information for different, goal-directed, purposes. The occipitotemporal pathway projects from primary visual cortex to the ventral areas of the brain in the temporal cortex and the occipitoparietal pathway projects to the dorsal areas of the brain to the parietal cortex and there is thought to be limited crosstalk between these streams (Ungerleider and Haxby, 1994; Young, 1992). Functionally, these areas are specified as being involved in either the recognition of objects (i.e. "what") or the perception of space for localisation ("where") or action ("how"). Evidence for this structural and functional dichotomy has been provided through lesion studies in animals (Mishkin et al., 1983) and from neuropsychological patient studies (Goodale and Milner, 1992). Furthermore, it is thought that this dual processing in the visual system is optimal to allow for the efficient processing of visual information for the purposes of recognition or action (e.g. Young, 1992).

If the visual system has an underlying structure that facilitates the efficient processing of information for perception, then we might ask whether the other senses are also similarly organised. Recent neuroimaging studies have suggested that this dual processing may also apply to both the human auditory system (Romanski et al., 1999) and the tactile system (Reed et al., 2005; Van Boven et al., 2005). In particular, Reed et al. conducted a neuroimaging study of tactile processing in which they asked participants to conduct either an object recognition or an object localisation task using the same, everyday familiar objects as stimuli and the same motor movements across conditions. They reported that these tasks selectively activated different brain regions. In particular brain regions involved in feature integration, such as inferior parietal areas, were activated during object recognition whereas brain regions involved in spatial processing, such as superior parietal regions, were activated during the object location task. Interestingly, a brain area typically involved in the recognition of familiar objects through vision (Grill-Spector et al., 1999; Malach et al., 1995) and touch (Amedi et al., 2001), that is, the lateral occipital complex, was not selectively activated during the object recognition task, suggesting that cortical activation of brain regions involved in object recognition may also depend on the nature of the exploration (i.e. recognition based on global shape or collection of local features) or on the possible role of imagery especially in the recognition of more familiar objects (see Lacey et al., 2009).

In a related study, we investigated brain activation during a tactile shape-matching and feature location task using the same unfamiliar objects interchangeably across these tasks (Newell et al., 2010b). See Fig. 14.1a for an example of some of our stimuli. Our results corroborate those reported by Reed et al. and suggest that, similar to the visual system, the tactile system is organised around distinct functional regions of the brain, each selectively involved in the processing of object information for recognition and for spatial localisation. Specifically, we found activation in ventral regions of the cortex, corresponding to the areas within and around the right lateral occipital complex, which were activated during the shape-matching task. In contrast, brain regions activated during the feature location task included more dorsal areas such as the left supramarginal and angular gyrus and also more ventral areas such as areas in and around the middle to lateral temporal region. Figure 14.1b provides an illustration of these activations.

Taken together, both the Reed et al. study and our study provide evidence for distinct processing of object shape and spatial location in the tactile system. In particular, clear associated areas of activation, particularly in the lateral occipital complex, are observed for object recognition tasks across both vision and touch (see also Amedi et al. 2001; Lacey et al., 2009). That the same cortical substrates underpin both visual and tactile object recognition suggests that similar object information is extracted by both modalities and consequently accessible to each of these senses. For the spatial tasks, although there is no evidence that a single cortical area was consistently activated by tactile and visual spatial perception across these studies, areas which were activated tended to lie within dorsal regions. Thus, the results from these studies suggest that the functional distinction between the object recognition and spatial tasks is preserved in the tactile system. The fact that specific cortical areas were not commonly activated across the studies may reflect task- and/or stimulus-specific spatial information processing. Nevertheless, the lack of

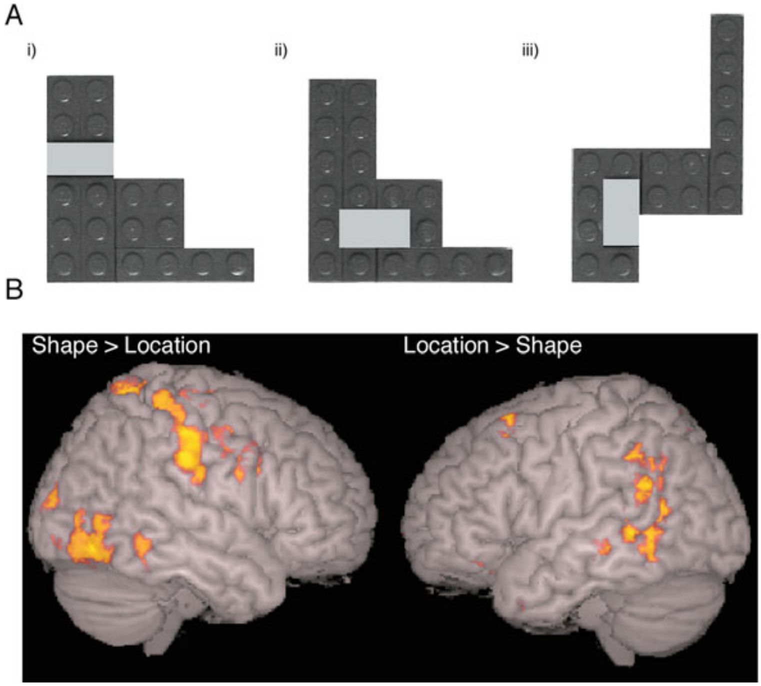

Fig. 14.1 (**a**) An example of the unfamiliar objects used in the Newell et al. (2010a, b) and Chan and Newell (2008) studies. Objects (i) and (ii) represent identically shaped objects, but the location of a feature on the objects changed across object shapes (i.e. same shape but different feature location). Objects (i) and (iii) represent differently shaped objects. (**b**) An illustration of the differential activations for each of the tasks ($z > 2.3$, $p < 0.05$ corrected)

common areas involved in visual and tactile spatial tasks raises the question of how efficiently spatial information is integrated across vision and touch since it suggests that common spatial information may not be extracted by these modalities. This question will be addressed in the following sections.

Recent behavioural evidence has corroborated the findings from our neuroimaging studies into the dual processing of object recognition and spatial localisation within and across the visual and tactile senses. We investigated the extent to which object information is processed separately and independently for the purpose of recognition and spatial location within vision and touch and also across these modalities using a dual-interference task (Chan and Newell, 2008). Using the same set of novel object stimuli interchangeably across tasks (see Fig. 14.1a), participants were required to perform an object shape or a feature location delayed match-to-sample task. In the shape task, for example, participants were presented with an unfamiliar object shape which they learned either through vision or touch. Following an inter-stimulus interval (ISI) of about 20 s, they were then presented with a second object shape which they had to judge as either the same or different to the shape of the first stimulus. The feature localisation task followed the same protocol but participants had to judge whether an object feature remained in the same relative location across the object surfaces or not. We then embedded a dual-interference, or secondary, task into these primary tasks by presenting a second matching task during the ISI of the primary task. This secondary task was either the same as the primary task (but with different stimuli) or the opposite task. In other words, if participants conducted a shape-based primary task then the embedded secondary task was either another shape-based task or a feature localisation task. Our aim was to investigate whether visual and haptic memory performance during a shape-matching task was interfered by another shape-related task or by a spatial task or by both, and whether visual and haptic performance on an object localisation task was affected by either an interfering shape task or a spatial task. We found that performance on a within-modal visual and haptic object shape-matching task was affected by a secondary shape-related task only and not a spatial task. Furthermore, our results suggested a double dissociation of task function since performance on a spatial task was affected by a secondary spatial task only and not a shape-matching task. Thus, our results suggest that there is a functional independence within both the visual system and the tactile system for the processing of object information for the purpose of recognition or spatial location. In our final experiment in that series, we investigated whether task-related interference was modality specific or independent of modality by embedding a secondary task that was conducted in a different modality to the primary task. Our results suggested that performance on spatial location tasks was affected by a crossmodal spatial task but not by a crossmodal shape-related task. In contrast, however, performance on a shape-matching task was affected by both shape and feature localisation secondary tasks. Thus, whereas a primary spatial task was not affected by a secondary, crossmodal shape-related task, a primary shape task was affected by both a crossmodal shape and a crossmodal spatial task.

Interestingly, this latter finding is consistent with our neuroimaging study discussed earlier in which we found that an object localisation task not only activates more dorsal areas in the parietal lobe but also areas in the temporal lobe which have been previously associated with shape perception, particularly the middle to lateral occipital areas (see Location > Shape activation image in Fig. 14.1b). In sum, evidence from both neuroimaging and behavioural studies converge to support the idea for functional independence of task-based information processing related to recognition and localisation not only within the visual and tactile modalities but also across these modalities, with the caveat that there may be some sharing of resources for object-based, feature localisation tasks across modalities. Since it has previously been argued that the functional distinction between “what” and “where” streams facilitates efficient information processing for recognition or action within the visual system (Young, 1992), we can probably assume that the distinction in the somatosensory system between cortical areas involved in object recognition or spatial localisation similarly benefits tactile perception. However, in order to maintain a coherent, multisensory perception of our world the brain must also allow for efficient cross-sensory information processing for the purpose of the task at hand. This may either be achieved by allowing the most appropriate sensory modality to dominate the perceptual outcome (Welch and Warren, 1986) or by merging the sensory information into a robust representation for perception or action (e.g. Ernst and Bülthoff, 2004). The following sections will examine the evidence supporting the idea that information is shared across vision and touch for the purpose of object recognition and spatial perception. Based on this evidence, I will argue that efficient crossmodal interactions seem to be determined by the extent to which principles of information processing are shared across vision and touch. Moreover, evidence is also emerging from neuroimaging studies that the cortical areas subserving object recognition and object localisation, although distinct within each modality, are largely overlapping across these modalities.

14.3 Evidence for Common Principles of Information Processing Across Vision and Touch for Object Recognition

Up to relatively recently, very little was understood about how we seem to effortlessly recognise objects under different ambient conditions such as changes in illumination, colour, viewpoint, and non-rigid changes due to movement. Although many researchers have investigated how object recognition is achieved by the visual system, a particular focus of this research has been on how vision solves the problem of object constancy, that is, how objects are recognised independent of viewpoint. This research effort was particularly driven by a need for computer scientists and engineers to develop systems of object recognition that could be adaptable for the purpose of automated recognition for manufacturing, security, and medical purposes. Much effort was conducted in the 1980s and 1990s on understanding how the human brain solves such an intractable problem as object constancy, with the idea that once this is understood in humans then these principles can be adopted into the design and development of computer systems and robots that could recognise objects at least well as humans, if not better. Needless to say, technology still lags far behind the capabilities of the human perceptual system although significant advances have been made through experimental psychology, cognitive neuroscience, and physiology into our understanding of how object constancy is achieved in the visual system (e.g. Biederman, 1987; Tarr and Bülthoff, 1998; Ullman, 1998).

14.3.1 Shape Constancy and the Visual Recognition of Objects Across Changes in Viewpoint

To account for view-invariant object recognition, many visual theorists initially proposed that objects were represented as structural descriptions in memory (e.g. Biederman, 1987; Marr, 1982). According to this proposal, the image of an object is deconstructed into its component 3D parts and it is the unique structural specification of these parts and their relative positions which allows the object to

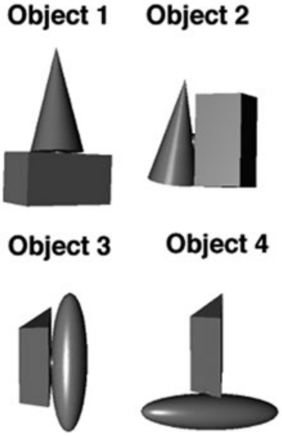

Fig. 14.2 An illustration of four distinct objects defined by either unique parts or unique arrangement of their parts. Objects 1 and 2 differ in the structural arrangements of their parts, as do objects 3 and 4. Objects 1 and 3, for example, differ both on their component parts and the structural arrangements of these parts

be recognised. Object constancy is effectively achieved provided the object parts, and the relations between them, can be resolved from the image of the object (e.g. Biederman and Gerhardstein, 1993). For example, Fig. 14.2 illustrates different objects consisting of same or different parts and arrangements of parts. All objects are recognised as different objects even though some share part shapes (e.g. Objects 1 and 2 comprise different arrangements of the same parts as do Objects 3 and 4). By reducing shapes to their basic component parts and specifying the arrangement of these parts, then object recognition should be efficient and, moreover, independent of viewpoint.

Biederman and colleagues provided evidence in support of this structural descriptions approach. They reported, for example, that recognition performance of images of objects from which their parts can be resolved is largely independent of viewpoint (Biederman and Gerhardstein, 1993) but when part information cannot be resolved then recognition is impaired (Biederman and Ju, 1998). However, other studies found evidence that object recognition is not independent of viewpoint and that even slight changes in the view of an object can disrupt performance (Bülthoff and Edelman, 1992; Newell and Findlay, 1997; Tarr, 1995; Tarr and Bülthoff, 1998). These findings led many researchers to propose that objects are represented in memory as a collection of select views and that recognition is consequently most efficient to views that match these stored views or the nearest stored view. This model was referred to as the "multiple views" model (e.g. Tarr and Bülthoff, 1998). Studies showing view dependency in object recognition tended to involve highly similar and unfamiliar object shapes as stimuli suggesting that the task demands may affect view-dependent performance (see, e.g. Newell, 1998; Hayward & Williams, 2000). Nevertheless, the recognition of highly familiar objects also seemed to be more efficient from some but not all views, known as canonical views (e.g. Palmer et al., 1981).

Thus it appeared that whereas the structural descriptions approach could account for the recognition of objects from different basic level categories, it could not account for the view-dependent recognition of novel, similar objects. On the other hand, proponents of the multiple views approach largely ignored the fact that objects can be compared based on their structure, suggesting that structural descriptions can be readily formed for the purpose of object perception. For example, in Fig. 14.2, objects 2 and 3 may be considered similar because of their basic part structure of one part adjacent to another. Similarly, objects 1 and 4 may be considered similar because of their part arrangements. If objects were stored as collection of views, then it would be difficult to explain how such structural comparisons could be achieved from an image description.

To account for the limitations of these approaches, Newell et al. (2005) proposed a hybrid model of object recognition. In a series of experiments, we found evidence in support of the idea that objects are represented as image-based parts but where the relative spatial locations of these parts are specified. Thus, our model took into account the idea that many objects can be deconstructed into component shapes and, moreover, that recognition can often be achieved when some of these components are obscured (such as a handle of a mug positioned at the back of the mug or a handbag positioned on its side). Since this model assumes that object parts are stored as images then it also predicts that recognition would be view dependent when these image-based parts are not presented in familiar or canonical orientations.

In sum, there is strong evidence from behavioural studies to suggest that the visual system does not solve the object constancy problem completely and that recognition is not completely invariant to changes in viewpoint. This evidence is also supported by studies investigating the neural correlates of object recognition which report view-dependent activations at the level of the single neuron in ventral areas of macaque brain (Logothetis and Pauls, 1995) and in the BOLD response in ventral cortical areas of the human brain (Andresen et al., 2009; Grill-Spector et al. 1998). However, other studies reported view invariance in the ventral stream with changes in rotation specifically affecting activation in more dorsal areas, particularly the IPS, more related to perception for action (e.g. James et al., 2002) or view invariance in neurons within later visual areas of the medial temporal lobe (Quiroga et al., 2005) possibly involving a consolidation of view-dependent activity from earlier visual areas. Given that the results from behavioural and neurophysiological studies converge to support the idea that recognition may not be invariant to viewpoint, it may thus seem like a conundrum why our everyday perception of objects is seemingly so robust and efficient. One potential solution may be that recognition occurs using a network of brain areas that act in unison to overcome the viewpoint problem either via a distributed coding of object information (e.g. Haxby et al., 2001) or with sparse, image-based representations (e.g. Reddy and Kanwisher, 2006) coupled with spatial processes such as mental rotation (e.g. Schendan and Stern, 2007; 2008). However, the fact that object recognition in the real world is not confined to one sensory modality may offer a clue as to how object constancy is achieved in the brain in that it may be via a combination of sensory information that objects are recognised most efficiently.

14.3.2 Shape Constancy and the Recognition of Static Objects in Vision and Touch

Although efforts in understanding how objects are recognised have mainly been concentrated on visual processing, object recognition is clearly not confined to the visual sense. Although objects can be identified through, for example, their characteristic sounds (e.g. the roar of a lion) or smells (e.g. a freshly peeled orange), it is only through the visual and tactile systems that object shape can be determined. Haptic perception of object shape can be very efficient and many studies have shown that familiar objects can be easily recognised using touch only (Gibson, 1962; Klatzky et al., 1985), even with very limited exposure to the object (Klatzky and Lederman, 1995). Since object information can be perceived through both vision and touch then it is possible that redundant shape information encoded across the senses may offer in a more robust representation of the object in memory (Ernst and Bülthoff, 2004).

If this is indeed the case, then it begs the question as to whether or not object recognition through the tactile system can provide the key to solving the object constancy problem. With this in mind, we investigated whether view-dependent object recognition was specific to visual processing and whether tactile object recognition is invariant to changes in object position (Newell et al., 2001). In agreement with Berkeley's statement that a blind man "By the Motion of his Hand he might discern the Situation of any Tangible Object placed within his Reach" we reasoned that since the hand is free to explore all surfaces of a 3D object (within certain size constraints), unlike vision which is constrained by optics, then haptic recognition should not necessarily be dependent on the object's position in the hand (i.e. its "view"). Surprisingly, we found that the haptic recognition of unfamiliar objects was dependent on the view of the object presented and that this cost in haptic recognition performance to a change in object view was similar to that found in visual object performance. Similar to the visual recognition of familiar objects, we also reported that the haptic system recognises some views of familiar objects more efficiently than other views (Woods et al., 2009).

However, in the Newell et al. (2001) study we found that the object views which promoted the most efficient recognition performance differed across modalities: whereas visual recognition was best for familiar frontal views of objects, the haptic system seemed to recognise better the back views of objects (with reference to the direction the observer is facing). Since each sensory system recognises different views most efficiently, then it seemed vision and haptics do not provide redundant information about object shape for the purpose of recognition. How then would object constancy be maintained if each system processes object information in a qualitatively different way? Although this seems impossible, the results of our study suggest a means by which object constancy is achieved across the senses: since vision and haptics encode different aspects of the object's shape then the combination of this non-redundant but complementary object information should result in a rich description of an object in memory that would help maintain object constancy over changes in object view. Such a rich representation should, therefore, mean that subsequent recognition of the object would be very efficient and indeed would likely not be dependent on viewpoint. In a recent study, Lacey et al. (2007) provided evidence to suggest that this is indeed the case. They investigated visual, haptic, and crossmodal object discrimination using a set of novel objects presented at different views rotated across one of three axes (*X*, *Y*, or *Z*). Their results corroborated previous findings that both visual and haptic object recognition is dependent on the view of the object but when combined they found that crossmodal object recognition performance is independent of viewpoint.

Both the Newell et al. (2001) and the Lacey et al. (2007) studies investigated visual and haptic object recognition across views of objects which were, by necessity, constrained in order to control the view information presented during training and test. In the real world, however, hand-sized objects are often picked up and palpated under free exploration. Indeed, in some previous studies on visual, haptic, and crossmodal recognition of familiar objects, haptic exploration was unconstrained whereas the object was presented in a fixed position for visual testing. These studies consistently reported that the sharing of object information across modalities is efficient (e.g. Easton et al. 1997; Reales and Ballesteros, 1999), suggesting that the information encoded by both modalities can be combined to allow for a rich, multisensory representation in memory. However, when familiar objects are used in a task, it is unclear the extent to which verbal labelling or other semantic information mediates crossmodal performance, especially when objects are presented in different views or positions across modalities. Indeed Berkeley hinted at the role which verbal labelling may play in crossmodal object recognition by stating that "Every Combination of Ideas is considered as one thing by the Mind, and in token thereof, is marked by one Name".

The use of unfamiliar objects that are not readily associated with distinct names gets around this issue. For example, Norman et al. (2004) used shapes based on natural objects from the same category (i.e. pepper or capsicum shapes) and found that crossmodal recognition performance was as good as within-modal performance. Norman et al. concluded that there are important similarities between vision and touch that allow for the same information to be represented in object memory (see also Gibson, 1979 for a similar conclusion). We also investigated whether the recognition of freely explored objects is efficient in vision and touch, and across these modalities, using the same set of novel objects as in our previous study (Ernst et al., 2007). Specifically, we tested both unimodal and bimodal (multisensory) recognition performance and found that unimodal performance was more efficient than bimodal performance. This finding seems to contradict the suggestions from earlier studies that vision and touch provide complementary information about an object which ultimately leads to a rich representation of the object in memory. However, close video analysis of the exploration procedures adopted during haptic exploration suggested that information encoding was optimised for efficient within-modal performance but not for crossmodal or bimodal performance.

The results from studies based on unconstrained exploration of objects suggest that object familiarity has no effect on the degree to which object information can be shared across the senses. Moreover, investigations into the recognition of static objects suggest that object constancy can be achieved by combining the inputs from both vision and touch into a rich representation of the object in memory.

14.3.3 Shape Constancy and the Recognition of Dynamic Objects in Vision and Touch

Whilst most investigations into object constancy have centred on the issue of viewpoint dependency, object shape can also be deformed by motion and any recognition system would need to account for how object constancy is maintained despite such changes to the shape. For example, in the animal world, shape can dramatically change whilst the animal is in motion and, moreover, this information may differ from when the animal is stationary and at rest. Furthermore, many small artefact objects can also change shape as a result of object motion: e.g. the opening of a book, mobile phone, or Swiss Army knife; or the rotation of a scissors blade; or the flipping of a stapler, can result in overall shape changes or reveal shape features not otherwise present in the object's image when it is closed. The shape information of these types of objects can change dramatically from one moment to the next but, as with changes in object view, we nevertheless seem to maintain object constancy with seemingly little effort.

Previous investigations on the visual recognition of dynamic objects have suggested that rigid motion information when combined with shape information can offer a unique cue to the identity of the object (e.g. Stone, 1998, 1999; Vuong and Tarr, 2004). For example, we found that object shapes associated with a particular movement pattern during learning were subsequently recognised more efficiently when shown with the same motion pattern than when shown moving in a different way (Newell et al., 2004). More recently, Setti and Newell (2009) reported that the visual recognition of unfamiliar objects in which non-rigid shape changes occur during motion is also affected by a change in the motion pattern of the parts of the objects. Thus, the findings of recent studies on the recognition of dynamic objects suggest that motion information can play an important role and that, moreover, objects may be stored in memory as spatiotemporal representations rather than static images of objects.

The tactile system obviously can play an important role in the perception of moving objects since many artefacts move as a consequence of haptic interactions. For example, a scissors changes shape as a consequence of movement of the hand. This begs the question, therefore, whether or not object motion affects the recognition of objects encoded through touch in the same way as those encoded through vision. Since motion is a cue for recognition in the visual domain then we can also ask whether the manner in which an object moves is a useful cue for recognition in the haptic domain. To that end we recently conducted a series of behavioural studies in which we investigated the role of motion information on the visual, haptic, and crossmodal recognition of object shapes (Whitaker et al., in prep.). We first created a set of unfamiliar objects, each with a moveable part which could rotate, flip, or slide on the object. At the beginning of the experiment, participants first learned a set of target objects either through vision only (by observing the object being moved by the experimenter) or through touch only (by actively palpating the object and moving the object part). Our results suggest that object motion is an important cue for recognition through touch: a moving target object was recognised better than its static counterpart. Furthermore, we found that both within-modal visual and haptic recognition benefited from the presence of the motion cue and the movement of the target objects also facilitated crossmodal recognition.

In order to assess whether motion is indeed an important cue for object recognition that is shared across modalities, we conducted a neuroimaging experiment to elucidate the neural correlates of crossmodal recognition of dynamic objects (Chan et al., 2010). We were specifically interested in investigating whether cortical areas known to be involved in visual motion and the visual recognition of dynamic objects were also involved in the haptic recognition of moving objects. Previous studies have found that area MT/MST is activated by dynamic information (e.g. Tootell et al., 1995) and also that motion implied in a static image is sufficient to activate this area (Kourtzi and Kanwisher, 2000). We first trained a group of participants to recognise a set of unfamiliar, moveable, and static objects using either vision or touch. We then presented static visual images of these objects to the participants whilst we recorded brain activations using fMRI. We found that area MT/MST was activated to images of objects previously learned as moving but not to objects learned as static. Surprisingly, this activation occurred to images of objects which were previously learned using either vision or touch (see Fig. 14.3). In other words, area MT/MST was active to both within modality and crossmodal presentations of previously learned dynamic objects.

These findings, together with the behavioural results reported earlier, suggest that both vision and touch contribute to the perception of moving objects and that, as such, both modalities may combine and share information in order to maintain object constancy not just in situations which involve changes in viewpoint but also those in which movement changes the shape information. Moreover, these findings are in contrast with Berkeley's conclusions on whether motion information is shared across modalities: he asserted that "… it clearly follows, that Motion perceivable by Sight is of a Sort distinct from Motion perceivable by Touch". However, although Berkeley did concede that for visuo-tactile perception "The Consideration of Motion, may furnish a new Field for Inquiry" it is perhaps surprising to note that this particular field remains relatively new three centuries later!

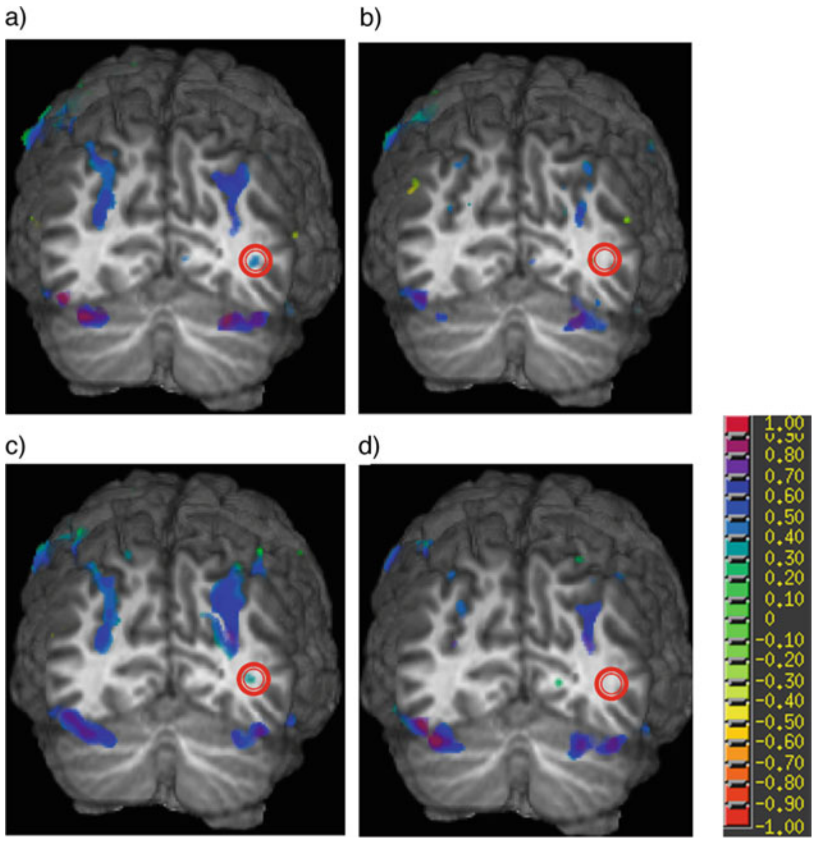

Fig. 14.3 Plot showing activation patterns across the four different learning conditions: (**a**) haptic motion; (**b**) haptic static; (**c**) visual motion and (**d**) visual static. Colours used represent positive and negative percentage change in the BOLD response (see key for colour coding). Area MT/MST (MT+) is highlighted on each map by a *red circle*

14.4 Evidence for Common Information Processing of “Where” Information Across Vision and Touch

14.4.1 Visual and Haptic Spatial Updating of Scenes

As previously mentioned, one of the greatest achievements of the human brain is the ability to perceive objects as constant despite dramatic changes in the projected retinal image, or in the tactile impressions, with changes in object position or shape deformities due to motion. An example of object constancy in the real world is that object recognition appears to remain invariant whilst we move around our environment even though the consequent changes in visual object information can be dramatic. For example, as we walk around a desk, the projected retinal image of the objects on that table can differ greatly depending on whether we have walked behind or in front of the desk. Although changes in object viewpoint consequently occur with observer motion, recognition performance does not seem to be affected in this situation. On the contrary, the recognition of views of an object that occur with observer motion is more efficient than the recognition of those same views when presented to a passive observer. Simons and his colleagues have attempted to account for this invariant object recognition with observer motion by proposing that extra-retinal information, such as vestibular or proprioception information, can inform the visual system of movement and consequently update the representation of the object in visual memory (e.g. Simons et al., 2002). Thus, information from other sensory modalities can update the representation of the object in visual memory to compensate for the consequent change in the visual projection of the object's image.

The finding that extra-retinal cues can update object representations in memory also pertains to the recognition of arrays of objects or scenes (Wang and Simons, 1999). However, up until recently, very little was known as to whether haptic representations are also updated with observer motion. Indeed it would seem that if object information is shared or accessible across modalities, as is suggested by research discussed previously, then spatial updating should be a process common to both vision and touch if a coherent perception of our world is to be achieved. In a series of studies, we investigated first whether the processes involved in the haptic recognition of object scenes is similar to the processes involved in the visual recognition of these scenes such that spatial information about object locations in a scene can be shared across modalities. Our previous research suggested that both "what" and "where" information is shared across modalities and, moreover, similar cortical substrates underpin these processes across modalities (Chan and Newell, 2008; Newell et al., 2010b). Similar to previous reports on the view-dependent recognition of scenes of objects in visual perception (Diwadkar and McNamara, 1997), we established that scene perception is also view dependent in haptic recognition in that the recognition of rotated scenes is more error prone than the recognition of scenes from a familiar view (Newell et al., 2005). Indeed, we recently found similar effects of scene rotation on haptic recognition using novel objects as we previously found using familiar shapes (see Fig. 14.4). However, in our study involving scenes of familiar shapes, we found that crossmodal recognition was less efficient than within-modal performance, although it was nevertheless better than chance. This cost in performance when crossing modalities did not seem to be due to differences in encoding across vision and touch (i.e. that vision can encode an object array in parallel, or from a single glance (Biederman et al., 1974; Rousselet et al., 2002) whereas haptics requires serial encoding of object positions). Instead, we argued that whilst some information can be shared across modalities, other spatial information is more modality specific and does not readily transfer across the senses.

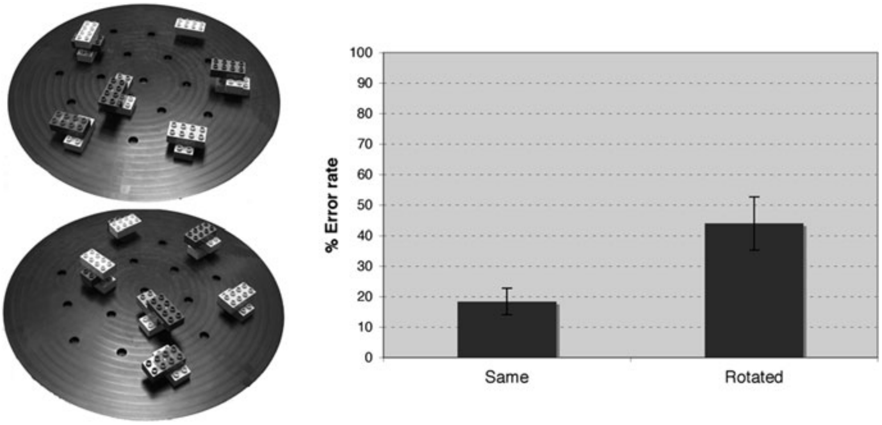

Fig. 14.4 An example of a scene of novel objects which participants were required to learn using haptics only (*top left* of image). An example of a test scene is shown in *bottom left* scene where an object has been displaced and the entire scene rotated by 60°. The plot shown on the *right* is the mean percentage errors made in novel scene recognition across non-rotated and rotated scenes

14.4.2 Crossmodal Updating in Scene Perception

The finding that scene recognition is better within than across modalities then led us to ask whether or not both visual and haptic scene perception benefit from spatial updating with observer movement. We found evidence to suggest that this is the case: the cost in both the visual and the haptic scene recognition performance as a consequence of passive rotation of the scene was prevented when the change in viewpoint was induced by observer motion (Pasqualotto et al., 2005). In particular, we replicated the results found in previous studies that the visual recognition of an array of familiar shapes is impaired when the scene is rotated relative to the observer and that this cost in recognition performance is removed when the rotated view is induced by the observer moving to a new position. We also extended this finding to the haptic domain and found that haptic scene representations are also updated with observer movement. The next question we asked was whether or not the representations of objects scenes are shared across modalities. In our most recent experiments, participants were required to learn the position of objects in a scene using either touch or vision only and we then tested their recognition in the other modality (Newell et al., 2010a). Prior to recognition the experimenter displaced one of the objects in the scene and the participant had to indicate which object had changed position. Furthermore, between learning and test the scene was either rotated relative to the passive observer, or the participant changed position. We found a cost in crossmodal performance when the scene was passively rotated. However, this cost was significantly reduced when the change in scene view was caused by a change in the observer's position. In other words, the visual or haptic representation of the object scene was updated during observer motion and this updating resulted in a benefit in recognition performance across modalities.

Since observer motion can update the representation of an object in memory, such that recognition in all modalities is benefited, this begs the question as to what mediates this crosstalk between the senses for the purpose of updating spatial representations. Previous studies have found that vision provides precision in perceptual decisions involving spatial information, even if those decisions are based on information encoded from another modality. For example, Newport et al. (2002) found that even visual information which was noninformative to a haptic task improves haptic performance.[1] Vision had a particular benefit when participants were encouraged to use a more allocentric than egocentric reference frame when performing the haptic task. More recently, Kappers, Postma, and colleagues further investigated the role of noninformative visual information on performance in a haptic parallel-matching task and found that vision affects the type of reference frame (i.e. from egocentric to allocentric) used to encode the haptic stimuli (Postma et al., 2008; Volcic et al., 2008). Moreover, they found that visual interfering information presented during the haptic task resulted in a cost in the haptic performance. Kaas et al. (2007) also found that noninformative visual information that is incongruent to haptic information can affect haptic performance on a parallelity task but only if the haptic information is encoded relative to an allocentric rather than an egocentric reference frame. These studies suggest that vision has a direct effect on haptic processing of spatial information by providing an allocentric reference frame to which haptic information is encoded.

14.4.3 The Role of Noninformative Visual Information on Haptic Scene Perception

We recently investigated the role of noninformative visual information on memory for object scenes encoded through touch. In these experiments, participants learned and were tested on their recognition of a scene of familiar objects using touch only. In separate conditions, participants could either view their surroundings (the test scene was never seen) or they were blindfolded during the task. We found evidence that noninformative vision can improve the haptic recognition of a scene of familiar object shapes (Pasqualotto et al., in prep.). However, the availability of visual information (albeit noninformative to the task) did not reduce the cost of recognising these haptic scenes when rotated, suggesting that scenes of familiar objects are stored as egocentric representations in memory, irrespective of the availability of noninformative visual information or of the encoding modality (see also Diwadkar and McNamara, 1997; Newell et al., 2005). Using virtual scenes, we manipulated

---

[1] Non-informative visual information is information that would not, on its own, be sufficient to solve the task. For example, seeing the surrounding room but without seeing the test stimuli would be considered ‘non-informative’ visual information.

the type of ambient visual information available during the haptic task to investigate the precise nature of the visual information that improves haptic performance. Specifically, participants could either see a furnished room, an empty room, or an image of the furniture without the room context. We found that spatial information (i.e. the presence of the room), not object landmarks (i.e. furniture only), was necessary in the visual image to benefit haptic performance. In conclusion, these studies suggest that vision can provide the optimum reference frame for encoding and retrieving spatial information through other senses although this benefit seems to be context dependent and does not necessarily affect the reference frame relative to which haptic spatial information is represented in memory.

If vision affects haptic spatial perception and memory then we might ask whether spatial perception is compromised in persons without visual experience. Some recent studies (e.g. Pasqualotto and Newell, 2007; Postma et al., 2008) tested haptic recognition of scenes of objects or haptic orientation perception in congenitally blind, late blind, and sighted people. There seems to be a consistent finding that tactile spatial perception is compromised in individuals who have impaired visual abilities, particularly those who were blind from early on in the course of development. Indeed it is well known from neurodevelopmental studies that early visual experience is required for normal development of the visual system (see, e.g. Lewis and Maurer, 2005 for a review) and that late intervention in repairing visual abnormalities can have long-term detrimental effects on the development of visual processing. The behavioural findings from haptic spatial perception suggest that the absence of visual experience can also affect the development of efficient spatial processing in another modality, namely touch. This finding has led some researchers to suggest that vision is the spatial sense which calibrates or modulates spatial perception in other, less spatially precise modalities (Thinus-Blanc and Gaunet, 1997).

14.5 Conclusions and Future Directions

The results of studies discussed above largely contradict Berkeley's assumption that "The Extension, Figures, and Motions perceived by Sight are specifically Distinct from the Ideas of Touch . . . nor is there any such thing as an Idea, or kind of Idea common to both Senses". On the contrary, evidence from the literature suggests that the manner in which object information is processed does not depend on the encoding modality. Moreover, both the visual and the tactile processing of object information seem to be underpinned by shared neural substrates. Since principles of information processing and neural resources are, to a large extent, shared across modalities for object recognition and localisation, this suggests that unisensory information is pooled together at some stage in perceptual processing. Although the time course of visuo-tactile interactions in the brain for the purpose of object perception has yet to be elucidated these interactions could occur either later on, according to the purpose of the task, or earlier such that all information is encoded into a multisensory representation to which each modality has access. Research on audio-visual processing for object recognition suggests that these interactions occur earlier on in perceptual processing than previously thought (e.g. Molholm et al., 2004). In any case, there seems to be little evidence that the sharing of information across modalities requires a distinct and separate recoding process that allows for vision and touch to share information for object recognition and localisation.

Although Berkeley concluded in his essay that visual and tactile processing are independent, he also asserted that where associations do exist between vision and touch these associations are not innate but are, instead, arbitrary and built from experience with the world. Specifically he proposed that "this Naming and Combining together of Ideas is perfectly Arbitrary, and done by the Mind in such sort, as Experience shows it to be most convenient". Although many studies have now provided evidence that vision and touch can efficiently share "ideas" for the purpose of object recognition and spatial perception, the extent to which these ideas are innate or hard wired versus the extent to which the associations are acquired through experience is, as yet, undetermined. However, evidence from developmental studies suggests that experience may be required for efficient crossmodal interactions to occur. For example, although some studies have provided evidence for crossmodal shape perception in neonates (e.g. Meltzoff and Borton, 1979) others have found that this crossmodal performance is not very efficient (Sann and Streri, 2007). Furthermore, some studies have found that whereas adult perception of spatial characteristics of object shape is based on a statistically optimal integration of information across vision and touch (Ernst and Banks, 2002) there is no evidence for this optimal integration in young children and indeed it does not seem to emerge until later on in development (Gori et al., 2008). As such, these studies suggest that although the sensory systems seem to be hard wired from birth to share information, the precision and efficiency with which information is integrated across the senses in adult perception seems to be dependent on experience. As Berkeley himself stated, ". . . this Connexion with Tangible Ideas has been learnt, at our first Entrance into the World, and ever since, almost every Moment of our Lives, it has been occurring to our Thoughts, and fastening and striking deeper on our Minds". However, very little is known about how this developmental process occurs and, moreover, what factors influence the normal development of multisensory integration. Future research into these areas would be very useful not only in elucidating the neurodevelopmental processes of perception but also in offering potential rehabilitative procedures to restore sensory function in a damaged brain or to counteract sensory decline due to either trauma or normal ageing.

In sum, an essay penned by the philosopher, George Berkeley, 300 years ago tapped into issues that are still relevant in the field of perception today. Although recent research has addressed many of the questions raised in that essay, some important issues on the nature and development of integration across vision and touch remain outstanding. It is encouraging to note, though, how much progress has been made on elucidating the behavioural and neural correlates of multisensory recognition in the last couple of decades and we can look forward to providing further empirical evidence in response to Berkeley's musings on the nature and development of visuo-tactile interactions in the near future.

References

Amedi A, Malach R, Hendler T, Peled S, Zohary E (2001) Visuo-haptic, object-related activation in the ventral visual pathway. Nat Neurosci 4:324–330

Andresen DR, Vinberg J, Grill-Spector K (2009) The representation of object viewpoint in human visual cortex. Neuroimage 45(2):522–536

Berkeley G (1709) An essay towards a new theory of vision. Jeremy Pepyat Booksellers, Dublin, Ireland

Biederman I (1987) Recognition-by-components: a theory of human image understanding. Psychol Rev 94:115–147

Biederman I, Gerhardstein PC (1993) Recognizing depth-view, rotated objects: evidence and conditions for three-dimensional viewpoint invariance. J Exp Psychol Hum Percept Perform 19:1162–1182

Biederman I, Ju G (1988) Surface versus edge-based determinants of visual recognition. Cogn Psychol 20(1):38–64

Biederman I, Rabinowitz JC, Glass AL, Stacy EW Jr (1974) On the information extracted from a glance at a scene. J Exp Psychol 103(3):597–600

Bülthoff HH, Edelman S (1992) Psychophysical support for a two-dimensional view interpolation theory of object recognition. Proc Natl Acad Sci USA 89:60–64

Chan JS, Newell FN (2008) Behavioral evidence for task-dependent “what” versus “where” processing within and across modalities. Percept Psychophys 70(1):36–49

Chan JS, Whitaker TA, Simões-Franklin C, Garavan H, Newell FN (2010) Implied haptic object motion activates visual area MT/MST. Neuroimage, 49(2):1708–1716

Diwadkar VA, McNamara TP (1997) Viewpoint dependence in scene recognition. Psycholog Sci 8:302–307

Easton RD, Srinivas K, Greene AJ (1997) Do vision and haptics share common representations? Implicit and explicit memory within and between modalities. J Exp Psychol Learn Mem Cogn 23(1):153–163

Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415(6870):429–433

Ernst MO, Bülthoff HH (2004) Merging the senses into a robust percept. Trends Cogn Sci 8: 162–169

Ernst MO, Lange C, Newell FN (2007) Multisensory recognition of actively explored objects. Can J Exp Psychol 61(3):242–253

Gibson JJ (1962) Observations on active touch. Psycholog Rev 69:477–491

Goodale MA, Milner AD (1992) Separate visual pathways for perception and action. Trends Neurosci 15:20–25

Gori M, Del Viva M, Sandini G, Burr DC (2008) Young children do not integrate visual and haptic form information. Curr Biol 18(9):694–698

Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, Malach R (1999) Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron 24:187–203

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293: 2425–2430

Hayward WG, Williams P (2000) Viewpoint dependence and object discriminability. Psychol Sci 11:7–12

James TW Humphrey GK Gati JS Menon RS, Goodale MA (2002) Differential effects of viewpoint on object-driven activation in dorsal and ventral streams. Neuron 35:793–801

Kaas AL, van Mier HI, Lataster J, Fingal M, Sack AT (2007) The effect of visuo-haptic congruency on haptic spatial matching. Exp Brain Res 183(1):75–85

Klatzky RL, Lederman SJ (1995) Identifying objects from a haptic glance. Percept Psychophys 57(8):1111–1123

Klatzky RL, Lederman SJ, Metzger V (1985) Identifying objects by touch: an expert system. Percept Psychophys 37:299–302

Kourtzi Z, Kanwisher N (2000) Activation in human MT/MST by static images with implied motion. J Cogn Neurosci 12(1):48–55

Lacey S, Peters A, Sathian K (2007) Cross-modal object recognition is viewpoint-independent. PLoS ONE 2(9):e890

Lacey S, Tal N, Amedi A, Sathian K (2009) A putative model of multisensory object representation. Brain Topogr 21(3–4):269–274

Lewis TL, Maurer D (2005) Multiple sensitive periods in human visual development: evidence from visually deprived children. Dev Psychobiol 46(3):163–183

Logothetis NK, Pauls J (1995) Psychophysical and physiological evidence for viewer-centered object representations in the primate. Cereb Cortex 5(3):270–288

Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RBH (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc Natl Acad Sci USA 92:8135–8139

Marr D (1982) Vision. Freeman Publishers, San Francisco

Meltzoff AN, Borton RW (1979) Intermodal matching by human neonates. Nature 282:403–404

Mishkin M, Ungerleider LG, Macko KA (1983) Object vision and spatial vision: two cortical pathways. Trends Neurosci 6:414–417

Molholm S, Ritter W, Javitt DC, Foxe JJ (2004) Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. Cereb Cortex 4:452–465

Newell FN (1998) Stimulus context and view dependence in object recognition. Perception 27(1):47–68

Newell FN, Ernst MO, Tjan BS, Bülthoff HH (2001) Viewpoint dependence in visual and haptic object recognition. Psychol Sci 12(1):37–42

Newell FN, Findlay JM (1997) The effect of depth rotation on object identification. Perception 26(10):1231–1257

Newell FN, Finucane CM, Lisiecka D, Pasqualotto A, Vendrell I (2010a) Active perception allows for spatial updating of object locations across modalities. Submitted

Newell, Hansen, Steven, Calvert (2010b) Tactile discrimination and localisation of novel object features activates both ventral and dorsal streams. Manuscript in preparation

Newell FN, Sheppard DM, Edelman S, Shapiro KL (2005) The interaction of shape- and location-based priming in object categorisation: evidence for a hybrid "what + where" representation stage. Vision Res 45(16):2065–2080

Newell FN, Wallraven C, Huber S (2004) The role of characteristic motion in object categorization. J Vis 4(2):118–129

Newell FN, Woods AT, Mernagh M, Bülthoff HH (2005) Visual, haptic and crossmodal recognition of scenes. Exp Brain Res 161(2):233–242

Newport R, Rabb B, Jackson SR (2002) Noninformative vision improves haptic spatial perception. Curr Biol 12(19):1661–1664

Norman JF, Norman HF, Clayton AM, Lianekhammy J, Zielke G (2004) The visual and haptic perception of natural object shape. Percept Psychophys 66(2):342–351

Palmer S, Rosch E, Chase P (1981) Canonical perspective and the perception of objects. In: Long JB, Baddeley AD (eds) Attention and performance, vol. IX. Erlbaum, Hillsdale, NJ, pp 135–151

Pasqualotto A, Finucane CM, Newell FN (2005) Visual and haptic representations of scenes are updated with observer movement. Exp Brain Res 166(3–4):481–488

Pasqualotto A, Newell FN (2007) The role of visual experience on the representation and updating of novel haptic scenes. Brain Cogn 65(2):184–194

Postma A, Zuidhoek S, Noordzij ML, Kappers AM (2008) Keep an eye on your hands: on the role of visual mechanisms in processing of haptic space. Cogn Proc 9(1):63–68

Postma A, Zuidhoek S, Noordzij ML, Kappers AM (2008) Haptic orientation perception benefits from visual experience: evidence from early-blind, late-blind, and sighted people. Percept Psychophys 70(7):1197–1206

Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I (2005) Invariant visual representation by single neurons in the human brain. Nature 435:1102–1107

Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP (1999) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat Neurosci 2:1131–1136

Reales JM, Ballestersos S (1999) Implicit and explicit memory for visual and haptic objects: cross-modal priming depends on structural descriptions. J Exp Psychol Learn Mem Cogn 25_(3) 644–663

Reddy L, Kanwisher N (2006) Coding of visual objects in the ventral stream. Curr Opin Neurobiol 16:408–414

Reed CL, Klatzky RL, Halgren E (2005) What vs. where in touch: an fMRI study. Neuroimage 25:718–726

Rousselet GA, Fabre-Thorpe M, Thorpe SJ (2002) Parallel processing in high-level categorization of natural images. Nat Neurosci 5(7):629–630

Sann C, Streri A (2007) Perception of object shape and texture in human newborns: evidence from cross-modal transfer tasks. Dev Sci 10(3):399–410

Schendan HE, Stern CE (2007) Mental rotation and object categorization share a common network of prefrontal and dorsal and ventral regions of posterior cortex. Neuroimage 35(3):1264–1277

Schendan HE, Stern CE (2008) Where vision meets memory: prefrontal-posterior networks for visual object constancy during categorization and recognition. Cereb Cortex 18(7):1695–1711

Setti A, Newell FN (2009) The effect of body and part-based motion on the recognition of unfamiliar objects. Vis Cogn 18(3):456–480

Simons DJ, Wang RF, Roddenberry D (2002) Object recognition is mediated by extraretinal information. Percept Psychophys 64(4):521–530

Stone JV (1999) Object recognition: view-specificity and motion-specificity. Vis Res 39(24): 4032–4044

Stone JV (1998) Object recognition using spatiotemporal signatures. Vis Res 38(7):947–951

Tarr MJ (1995) Rotating objects to recognize them: a case study of the role of viewpoint dependency in the recognition of three-dimensional objects. Psychon Bull Rev 2:55–82

Tarr MJ, Bülthoff HH (1998) Image-based object recognition in man, monkey and machine. Cognition 67:1–20

Thinus-Blanc C, Gaunet F (1997) Representation of space in blind persons: vision as a spatial sense? Psychol Bull 121(1):20–42

Tootell RB, Reppas JB, Kwong KK, Malach R, Born RT, Brady TJ, Rosen BR, Belliveau JW (1995) Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. J Neurosci 15(4):3215–3230

Ungerleider LG, Haxby JV (1994) ‘What’ and ‘where’ in the human brain. Curr Opin Neurobiol 4, 157–165

Ullman S (1998) Three-dimensional object recognition based on the combination of views. Cogn 67:21–44

Van Boven RW, Ingeholm JE, Beauchamp MS, Bikle PC, Ungerleider LG (2005) Tactile form and location processing in the human brain. Proc Natl Acad Sci U S A 102(35):12601–12605

Volcic R, van Rheede JJ, Postma A, Kappers AM (2008) Differential effects of non-informative vision and visual interference on haptic spatial processing. Exp Brain Res 190(1):31–41

Vuong QC, Tarr MJ (2004) Rotation direction affects object recognition. Vis Res 44(14): 1717–1730

Welch RB, Warren DH (1986) Intersensory interactions. In: Boff KR, Kaufman L, Thomas JP (eds) Handbook of perception and performance, vol. 1. Sensory processes and perception. John Wiley and Sons, New York, pp 25–1–25–36

Whitaker TA, Chan J, Newell FN (2010) The role of characteristic motion in haptic and visuo-haptic object recognition. Manuscript in preparation

Wang RF, Simons DJ (1999) Active and passive scene recognition across views. Cognition 70(2):191–210

Woods AT, Moore A, Newell FN (2009) Canonical views in haptic object perception. Perception 37(12):1867–1878

Young MP (1992) Objective analysis of the topological organization of the primate cortical visual system. Nature 358(6382):152–155